

COMPUTING QUASISOLUTIONS OF NONLINEAR INVERSE PROBLEMS VIA EFFICIENT MINIMIZATION OF TRUST REGION PROBLEMS

Barbara Kaltenbacher and Franz Rendl and Elena Resmerita
Institute of Mathematics, Alpen-Adria-Universität Klagenfurt

September 11, 2015

In this paper we present a method for the regularized solution of nonlinear inverse problems, based on Ivanov regularization (also called method of quasi solutions or constrained least squares regularization). This leads to the minimization of a non-convex cost function under a norm constraint, where non-convexity is caused by nonlinearity of the inverse problem. Minimization is done by iterative approximation, using (non-convex) quadratic Taylor expansions of the cost function. This leads to repeated solution of quadratic trust region subproblems with possibly indefinite Hessian. Thus the key step of the method consists in application of an efficient method for solving such quadratic subproblems, developed by Rendl and Wolkowicz [10]. We here present a convergence analysis of the overall method as well as numerical experiments.

1 INTRODUCTION

Consider the nonlinear inverse problem of recovering $x \in X$ in

$$(1) \quad F(x) = y$$

– with a forward operator $F : \mathcal{D}(F) \subseteq X \rightarrow Y$ mapping from a Hilbert space X to a Banach space Y – from noisy measurements y^δ of $y \in Y$ satisfying

$$(2) \quad \mathcal{S}(y, y^\delta) \leq \delta$$

where $\mathcal{S} : Y^2 \rightarrow \mathbb{R}^+$ is a distance measure quantifying the data misfit, for instance a norm but possibly also a more involved expression such as the Kullback Leibler divergence arising from certain statistic measurement noise models.

The regularized solution of (1) via the method of quasi solutions (also called Ivanov regularization) leads to minimization problems of the form

$$(3) \quad x_\rho^\delta \in \operatorname{argmin}\{r^\delta(x) : x \in X, \|x\|^2 \leq \rho\},$$

where

$$r^\delta(x) = \mathcal{S}(F(x), y^\delta)$$

with an appropriately chosen radius ρ that plays the role of a regularization parameter here, cf., [2, 4, 5, 6, 8, 9, 11, 13]. The selection of ρ can be done in an a priori fashion if the norm of some exact solution x^\dagger to (1) is known. Namely, in that case the choice $\rho = \|x^\dagger\|^2$ can be shown to be optimal. Otherwise, also a noise level dependent a posteriori choice of δ according to Morozov's discrepancy principle, i.e., $\rho = \rho(\delta)$ such that

$$(4) \quad \delta < r^\delta(x_\rho) \leq \tau\delta$$

for some $\tau > 1$ fixed independent of the noise level δ , leads to convergence, cf. [1].

In this paper we wish to exploit the obvious relation to trust region subproblems to take advantage of an efficient algorithm proposed in [10] for solving quadratic trust region subproblems with not necessarily positive semidefinite Hessian – a situation that is highly relevant here in view of the fact that the cost functional in (3) exhibits potential nonconvexity due to nonlinearity of the forward operator F and/or the distance measure \mathcal{S} .

To this end, we first of all discretize the problem by restriction of minimization to finite dimensional subspaces $X_n \subseteq X$ and the use of computational approximations of the involved operators and norms, which leads to a sequence of finite dimensional problems

$$(5) \quad x_{n,\rho}^\delta \in \operatorname{argmin}\{r_n^\delta(x_n) : x_n \in X_n, \|x_n\|_n^2 \leq \rho\}$$

where

$$r_n^\delta(x_n) = \mathcal{S}_n(F_n(x_n), y^\delta),$$

where $\mathcal{S}_n : Y^2 \rightarrow \mathbb{R}$ $F_n : X_n \rightarrow Y$ are approximations, e.g., due to discretization, of \mathcal{S} and F , respectively.

We also consider the practically relevant situation that these minimization problems are not solved with infinite precision but in an inexact sense, i.e., we will use regularized approximations $\hat{x}_{n,\rho}^\delta \in X_n$ satisfying

$$(6) \quad \|\hat{x}_{n,\rho}^\delta\|_n^2 \leq \rho \text{ and } r_n^\delta(\hat{x}_{n,\rho}^\delta) \leq \min_{x_n \in X_n} \{r_n^\delta(x_n) : \|x_n\|_n^2 \leq \rho\} + \eta_n^\delta$$

and choose $\rho = \rho(\delta)$ such that

$$(7) \quad \delta + \underline{\eta}_n^\delta < r_n^\delta(\hat{x}_{n,\rho}^\delta) \leq \tau\delta + \bar{\eta}_n^\delta$$

with appropriately chosen tolerances η_n^δ , $\underline{\eta}_n^\delta$, $\bar{\eta}_n^\delta$. A first requirement on these tolerances to admit solutions to (6) and (7) is

$$(8) \quad \eta_n^\delta \geq 0 \text{ and } \underline{\eta}_n^\delta < (\tau - 1)\delta + \bar{\eta}_n^\delta$$

We mention in passing that this nonconvex case is also of special interest for Ivanov regularization since this is also the situation in which it is in general not equivalent to Tikhonov regularization.

2 REGULARIZING PROPERTY OF $\hat{x}_{n,\rho}^\delta$

To analyze convergence of $\hat{x}_{n,\rho}^\delta$ to an exact solution x^\dagger of (1) as $\delta \rightarrow 0$, we first of all state a straightforward monotonicity property for the exact minimizers at fixed discretization level n .

Lemma 2.1. *The mapping $\rho \mapsto r_n^\delta(x_{n,\rho}^\delta)$ with $x_{n,\rho}^\delta$ according to (5) is monotonically decreasing.*

Proof. Since the admissible set is larger for larger radius and we minimize the same cost function, the assertion is obvious. \square

Moreover, we get a uniform bound on the radii chosen according to (7) with $\hat{x}_{n,\rho}^\delta$ satisfying (6), provided the tolerances are chosen appropriately.

Lemma 2.2. *Let (8) and*

$$(9) \quad \underline{\eta}_n^\delta - \eta_n^\delta \geq r_n^\delta(P_n x^\dagger) - r^\delta(x^\dagger)$$

hold, where P_n is the orthogonal projection onto X_n .

Then $\rho(\delta)$ according to (7) satisfies the estimate

$$(10) \quad \rho(\delta) \leq \|P_n x^\dagger\|_n^2 =: \rho_n^\dagger$$

Proof. By (6) and (7), as well as (2) we have, for $\rho = \rho(\delta)$,

$$r_n^\delta(x_{n,\rho}^\delta) \geq r_n^\delta(\hat{x}_{n,\rho}^\delta) - \eta_n^\delta > \delta + \underline{\eta}_n^\delta - \eta_n^\delta \geq r^\delta(x^\dagger) + \underline{\eta}_n^\delta - \eta_n^\delta \geq r_n^\delta(P_n x^\dagger) \geq r_n^\delta(x_{n,\rho_n^\dagger}^\delta),$$

where in the last two inequalities we have used (9) and minimality of $x_{n,\rho_n^\dagger}^\delta$ (together with feasibility of $P_n x^\dagger$ for the discretized problem). Thus the assertion follows by contraposition in Lemma 2.1 \square

To prove convergence and convergence rates, we will make use of two general results Propositions 2.3 and 2.4, see also Theorems 2.4 and 2.7 in [1] for a slightly different setting. Here for some fixed maximal noise level $\bar{\delta}$, $(y^\delta)_{\delta \in (0, \bar{\delta}]}$ is a family of data satisfying (2) and $(\tilde{x}^\delta)_{\delta \in (0, \bar{\delta}]}$ a family of regularized approximations (defined by (6) with (7) or by some other regularization method).

Proposition 2.3. *Let F be weakly sequentially closed at y in the sense that*

$$(11) \quad \begin{aligned} &\forall (y_k)_{k \in \mathbb{N}}, (x_k)_{k \in \mathbb{N}} : \\ &(x_k \rightharpoonup x \text{ and } \mathcal{S}(F(x_k), y_k) \rightarrow 0 \text{ and } \mathcal{S}(y, y_k) \rightarrow 0 \Rightarrow y \in \mathcal{D}(F) \text{ and } F(x) = y \end{aligned}$$

and let uniform boundedness

$$(12) \quad \exists \bar{\rho} > 0 \quad \forall \delta \in (0, \bar{\delta}] : \quad \|\tilde{x}^\delta\|^2 \leq \bar{\rho}$$

as well as convergence of the operator values in the sense of the distance measure \mathcal{S}

$$(13) \quad \mathcal{S}(F(\tilde{x}^\delta), y^\delta) \rightarrow 0 \text{ as } \delta \rightarrow 0$$

hold.

Then there exists a weakly convergent subsequence of $(\tilde{x}^\delta)_{\delta \in (0, \bar{\delta}]}$ and the limit x^* of every weakly convergent subsequence of $(\tilde{x}^\delta)_{\delta \in (0, \bar{\delta}]}$ satisfies (1).

If for such a weakly convergent subsequence $(\tilde{x}^{\delta_k})_{k \in \mathbb{N}}$ with limit x^* , (12) can be strengthened to

$$(14) \quad \limsup_{k \rightarrow \infty} \|\tilde{x}^{\delta_k}\|^2 \leq \|x^*\|^2,$$

then we even have strong convergence $\tilde{x}^{\delta_k} \rightarrow x^*$ as $k \rightarrow \infty$.

If the solution x^\dagger to (1) is unique then $\tilde{x}^\delta \rightarrow x^\dagger$ as $\delta \rightarrow 0$ under condition (12) and $\tilde{x}^\delta \rightarrow x^\dagger$ as $\delta \rightarrow 0$ under condition $\limsup_{\delta \rightarrow 0} \|\tilde{x}^\delta\|^2 \leq \|x^\dagger\|^2$.

To state convergence with rates to some solution x^\dagger of (1) we make use of a variational source condition

$$(15) \quad \exists \beta \in (0, 1) \quad \forall x \in \mathcal{B}_R(x^\dagger) \quad 2(x^\dagger, x^\dagger - x) \leq \beta \|x^\dagger - x\|^2 + \varphi(\mathcal{S}(F(x^\dagger), F(x)))$$

with some radius $R > 0$ and some index function $\varphi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, (i.e. φ is monotonically increasing and $\lim_{t \rightarrow 0} \varphi(t) = 0$). Condition (15) is a condition on the smoothness of x^\dagger that is the stronger the faster φ decays to zero as $t \rightarrow 0$. It is, e.g., satisfied with $\mathcal{S}(y_1, y_2) = \frac{1}{2} \|y_1 - y_2\|^2$ and $\varphi(t) \sim t$ if x^\dagger lies in the range of the adjoint of the linearized forward operator (which is typically a smoothing operator) and F' is Lipschitz continuous

$$x^\dagger = F'(x^\dagger)^* w \text{ and } \forall x \in \mathcal{B}_R(x^\dagger) : \|F'(x) - F'(\tilde{x})\| \leq L \|x - \tilde{x}\| \text{ and } L \|w\| < 1.$$

By a homogeneity argument for the case of linear F it can be seen that the fastest possible decay of φ at zero that gives a reasonable assumption in (15) is $\varphi(t) \sim t$.

Proposition 2.4. *Let (15) hold and \tilde{x}^δ be contained in $\mathcal{B}_R(x^\dagger)$ for all $\delta \in (0, \bar{\delta}]$. Moreover, assume that there exist constants $C_1, C_2, C_3 > 0$ such that*

$$(16) \quad \forall \delta \in (0, \bar{\delta}] : \quad \|\tilde{x}^\delta\|^2 \leq \|x^\dagger\|^2 + C_1 \varphi(C_2 \delta) \text{ and } \mathcal{S}(F(\tilde{x}^\delta), y^\delta) \leq C_3 \delta.$$

and that \mathcal{S} satisfies the generalized triangle inequality

$$(17) \quad \mathcal{S}(y_1, y_2) \leq C_4 (\mathcal{S}(y_1, y_3) + \mathcal{S}(y_2, y_3))$$

for some $C_4 > 0$ and all $y_1, y_2 \in Y$.

Then the convergence rate

$$\|\tilde{x}^\delta - x^\dagger\|^2 = O(\varphi(C \delta)) \text{ as } \delta \rightarrow 0$$

holds with $C = \max\{C_2, C_4(C_3 + 1)\}$.

Proof. From (15) with $x = \tilde{x}^\delta$ and (16), as well as (17), (2), and monotonicity of φ we get

$$\begin{aligned}\|\tilde{x}^\delta - x^\dagger\|^2 &= \|\tilde{x}^\delta\|^2 - \|x^\dagger\|^2 + 2(x^\dagger, x^\dagger - \tilde{x}^\delta) \\ &\leq \beta \|\tilde{x}^\delta - x^\dagger\|^2 + C_1 \varphi(C_2 \delta) + \varphi(C_4(\mathcal{S}(F(\tilde{x}^\delta), y^\delta) + \delta)).\end{aligned}$$

□

Corollary 2.5. *Let (11) hold and let $\hat{x}_{n(\delta), \rho(\delta)}^\delta$ be defined by (6), (7), with the discretization level $n = n(\delta)$ and the tolerances $\eta_{n(\delta)}^\delta, \underline{\eta}_{n(\delta)}^\delta, \bar{\eta}_{n(\delta)}^\delta$ chosen such that (8), (9) and*

$$(18) \quad \left. \begin{aligned} &\|P_{n(\delta)} x^\dagger\|_{n(\delta)}^2 - \|x^\dagger\|^2 \leq C_{1,1} \varphi(C_2 \delta) \\ &\|x\|^2 - \|x\|_{n(\delta)}^2 \leq C_{1,2} \varphi(C_2 \delta) \\ &r^\delta(x) - r_{n(\delta)}^\delta(x) \leq \tau_1 \delta \end{aligned} \right\} \text{ for } x = \hat{x}_{n(\delta), \rho(\delta)}^\delta$$

$$\bar{\eta}_{n(\delta)}^\delta \leq \tau_2 \delta$$

hold for fixed constants $\tau_1, \tau_2, C_{1,1}, C_{1,2}, C_2 > 0$ independent of δ .

Then $\hat{x}_{n(\delta), \rho(\delta)}^\delta$ converges to x^\dagger subsequentially in the sense of Proposition 2.3.

If additionally a variational source condition (15) and the generalized triangle inequality (17) holds, then

$$\|\hat{x}_{n(\delta), \rho(\delta)}^\delta - x^\dagger\|^2 = O(\varphi(C\delta)) \text{ as } \delta \rightarrow 0,$$

with $C = \max\{C_2, C_4(\tau + \tau_1 + \tau_2 + 1)\}$.

Proof. By Lemma 2.2 and (18) we have

$$\|\hat{x}_{n(\delta), \rho(\delta)}^\delta\|^2 \leq \|x^\dagger\|^2 + (C_{1,1} + C_{1,2}) \varphi(C_2 \delta).$$

Moreover, by (7) and (18) we can estimate

$$\mathcal{S}(F(\hat{x}_{n(\delta), \rho(\delta)}^\delta), y^\delta) = r^\delta(\hat{x}_{n(\delta), \rho(\delta)}^\delta) \leq r_n^\delta(\hat{x}_{n(\delta), \rho(\delta)}^\delta) + \tau_1 \delta \leq (\tau + \tau_1 + \tau_2) \delta.$$

□

3 COMPUTATION OF $\hat{x}_{n, \rho}^\delta$ AND OF $\rho(\delta)$

3.1 A NEWTON TYPE ITERATION FOR COMPUTING $\hat{x}_{n, \rho}^\delta$

For fixed discretization and noise level n, δ and fixed radius $\rho < \rho(\delta)$, we approximate the nonlinear trust region subproblem (5)

$$(19) \quad \min_{x_n \in X_n} r_n^\delta(x_n) \text{ s.t. } \|x_n\|_n^2 \leq \rho$$

by a sequence of quadratic subproblems arising from second order Taylor expansion of the cost function

$$(20) \quad \min_{x_n \in X_n} q^k(x_n) \text{ s.t. } \|x_n\|_n^2 \leq \rho$$

with

$$(21) \quad q^k(x) = r_n^\delta(x^k) + r_n^{\delta'}(x^k)(x - x^k) + \frac{1}{2}r_n^{\delta''}(x^k)(x - x^k)^2$$

where x^k is some current iterate. Necessary second order optimality conditions for these two minimization problems (in case of (20) they will also be sufficient, cf. [12]) are existence of $\lambda_\rho, \lambda^{k+1} \in \mathbb{R}$ such that

$$(22) \quad r_n^{\delta'}(x_{n,\rho}^\delta) + \lambda_\rho \langle x_{n,\rho}^\delta, \cdot \rangle_n = 0$$

$$(23) \quad \lambda_\rho \geq 0, \quad \|x_{n,\rho}^\delta\|_n^2 \leq \rho, \quad \lambda_\rho(\|x_{n,\rho}^\delta\|_n^2 - \rho) = 0$$

$$(24) \quad \forall w \in C(x_{n,\rho}^\delta) : r_n^{\delta''}(x_{n,\rho}^\delta)w^2 \geq 0$$

for (19), where for these simple constraints the critical cone is given by

$$C(x_{n,\rho}^\delta) = \begin{cases} X_n & \text{if } \|x_{n,\rho}^\delta\|_n^2 < \rho \\ \{x_{n,\rho}^\delta\}^\leq = \{v \in X_n : \langle x_{n,\rho}^\delta, v \rangle_n \leq 0\} & \text{if } \|x_{n,\rho}^\delta\|_n^2 = \rho \text{ and } \lambda_\rho = 0 \\ \{x_{n,\rho}^\delta\}^\perp = \{v \in X_n : \langle x_{n,\rho}^\delta, v \rangle_n = 0\} & \text{if } \|x_{n,\rho}^\delta\|_n^2 = \rho \text{ and } \lambda_\rho > 0 \end{cases}$$

and

$$(25) \quad r_n^{\delta'}(x_n^k) + r_n^{\delta''}(x_n^k)(x_n^{k+1} - x_n^k) + \lambda^{k+1} \langle x_n^{k+1}, \cdot \rangle_n = 0$$

$$(26) \quad \lambda^{k+1} \geq 0, \quad \|x_n^{k+1}\|_n^2 \leq \rho, \quad \lambda^{k+1}(\|x_n^{k+1}\|_n^2 - \rho) = 0$$

$$(27) \quad \forall w \in X_n : r_n^{\delta''}(x_n^k)w^2 + \lambda^{k+1}\|w\|_n^2 \geq 0$$

for (20) with (21). For the error $x_n^{k+1} - x_{n,\rho}^\delta$ this implies

$$(28) \quad \begin{aligned} & (r_n^{\delta''}(x_{n,\rho}^\delta) + \lambda_\rho I_n)(x_n^{k+1} - x_{n,\rho}^\delta) \\ &= r_n^{\delta''}(x_{n,\rho}^\delta)(x_n^{k+1} - x_{n,\rho}^\delta) + \langle -\lambda_\rho x_{n,\rho}^\delta + \lambda^{k+1} x_n^{k+1} + (\lambda_\rho - \lambda^{k+1})x_n^{k+1}, \cdot \rangle_n \\ &= r_n^{\delta''}(x_{n,\rho}^\delta)(x_n^{k+1} - x_{n,\rho}^\delta) + r_n^{\delta'}(x_{n,\rho}^\delta) - r_n^{\delta'}(x_n^k) - r_n^{\delta''}(x_n^k)(x_n^{k+1} - x_n^k) + (\lambda_\rho - \lambda^{k+1}) \langle x_n^{k+1}, \cdot \rangle_n \\ &=: t_\rho + (\lambda_\rho - \lambda^{k+1}) \langle x_n^{k+1}, \cdot \rangle_n \end{aligned}$$

where $I_n : X_n^2 \rightarrow \mathbb{R}$, $I_n(x_n, \tilde{x}_n) = \langle x_n, \tilde{x}_n \rangle_n$, and

$$(29) \quad \begin{aligned} & (r_n^{\delta''}(x_n^k) + \lambda^{k+1} I_n)(x_n^{k+1} - x_{n,\rho}^\delta) \\ &= r_n^{\delta''}(x_n^k)(x_n^{k+1} - x_{n,\rho}^\delta) + \langle -\lambda_\rho x_{n,\rho}^\delta + \lambda^{k+1} x_n^{k+1} + (\lambda_\rho - \lambda^{k+1})x_{n,\rho}^\delta, \cdot \rangle_n \\ &= r_n^{\delta''}(x_n^k)(x_n^k - x_{n,\rho}^\delta) + r_n^{\delta'}(x_{n,\rho}^\delta) - r_n^{\delta'}(x_n^k) + (\lambda_\rho - \lambda^{k+1}) \langle x_{n,\rho}^\delta, \cdot \rangle_n \\ &=: t^{k+1} + (\lambda_\rho - \lambda^{k+1}) \langle x_{n,\rho}^\delta, \cdot \rangle_n \end{aligned}$$

where under a Lipschitz condition on $r_n^{\delta''}$

$$(30) \quad \forall x_n, \tilde{x}_n \in \mathcal{D}(F_n) : \|r_n^{\delta''}(x) - r_n^{\delta''}(\tilde{x})\|_n \leq L\|x - \tilde{x}\|_n$$

we have

$$\begin{aligned}
\|t_\rho\|_n &= \int_0^1 \left[r_n^{\delta''}(x_{n,\rho}^\delta + \theta(x_{n,\rho}^\delta - x_n^k)) - r_n^{\delta''}(x_{n,\rho}^\delta) \right] d\theta(x_{n,\rho}^\delta - x_n^k) \\
&\quad + \left[r_n^{\delta''}(x_{n,\rho}^\delta) - r_n^{\delta''}(x_n^k) \right] (x_n^{k+1} - x_n^k)_n \\
&\leq \frac{L}{2} \|x_{n,\rho}^\delta - x_n^k\|_n^2 + L \|x_{n,\rho}^\delta - x_n^k\|_n \|x_n^{k+1} - x_n^k\|_n \\
\|t^{k+1}\|_n &= \int_0^1 \left[r_n^{\delta''}(x_{n,\rho}^\delta + \theta(x_{n,\rho}^\delta - x_n^k)) - r_n^{\delta''}(x_{n,\rho}^\delta) \right] d\theta(x_{n,\rho}^\delta - x_n^k) \\
&\leq \frac{L}{2} \|x_{n,\rho}^\delta - x_n^k\|_n^2
\end{aligned}$$

Testing the sum of (28) and (29) with $x_n^{k+1} - x_{n,\rho}^\delta$ and using the fact that

$$\begin{aligned}
(\lambda_\rho - \lambda^{k+1}) \langle x_n^{k+1} + x_{n,\rho}^\delta, x_n^{k+1} - x_{n,\rho}^\delta \rangle_n &= (\lambda_\rho - \lambda^{k+1}) (\|x_n^{k+1}\|_n^2 - \|x_{n,\rho}^\delta\|_n^2) \\
&= \begin{cases} \lambda_\rho (\|x_n^{k+1}\|_n^2 - \rho) & \text{if } \|x_{n,\rho}^\delta\|_n^2 = \rho \\ \lambda^{k+1} (\|x_{n,\rho}^\delta\|_n^2 - \rho) & \text{if } \|x_{n,\rho}^\delta\|_n^2 < \rho \end{cases} \leq 0
\end{aligned}$$

(the latter representation is readily checked by a distinction of the cases $\|x_n^{k+1}\|_n^2 = / < \rho$) we end up with

$$\begin{aligned}
(31) \quad & (r_n^{\delta''}(x_{n,\rho}^\delta) + r_n^{\delta''}(x_n^k) + \lambda^{k+1} I_n) (x_n^{k+1} - x_{n,\rho}^\delta)^2 + \lambda_\rho \|x_n^{k+1} - x_{n,\rho}^\delta\|_n^2 \\
& \leq L (\|x_{n,\rho}^\delta - x_n^k\|_n^2 + \|x_{n,\rho}^\delta - x_n^k\|_n \|x_n^{k+1} - x_n^k\|_n) \|x_n^{k+1} - x_{n,\rho}^\delta\|_n.
\end{aligned}$$

From this we wish to extract an estimate on the error norm $\|x_n^{k+1} - x_{n,\rho}^\delta\|_n$. However, the operator $r_n^{\delta''}(x_{n,\rho}^\delta) + r_n^{\delta''}(x_n^k) + \lambda^{k+1} I_n$ on the left hand side in (31) is positive semidefinite only on the critical cone $C(x_{n,\rho}^\delta)$. Thus in the case $\lambda_\rho > 0$ in which $C(x_{n,\rho}^\delta)$ only consists of directions orthogonal to $x_{n,\rho}^\delta$, we have to make an additional assumption to avoid negative contributions on the left hand side (that generally might have even larger modulus than the good term $\lambda_\rho \|x_n^{k+1} - x_{n,\rho}^\delta\|_n^2$)

$$(32) \quad r_n^{\delta''}(x_{n,\rho}^\delta) + \lambda_\rho I_n \text{ is positive definite, where } \lambda_\rho = -\frac{(r_n^{\delta'}(x_{n,\rho}^\delta), x_{n,\rho}^\delta)}{\rho}.$$

Here we have used the fact that the Lagrange multiplier can be explicitly represented due to the necessary optimality conditions (22) (23).

Assumption (32) is obviously satisfied if the cost function r_n^δ is convex, but it also admits nonconvexity possibly arising due to nonlinearity of F and/or S , as the following example shows.

Example 1. $n = 1$, $x^\dagger > 0$, $C \in (\frac{1}{12x^{\dagger 2}}, \frac{1}{4x^{\dagger 2}})$, $D > \frac{48Cx^{\dagger 2}-1}{12(1/(2\sqrt{C})-x^\dagger)^2} > 0$

$$r_1^\delta(x) = \frac{1}{2}(x - x^\dagger)^2 - C(x - x^\dagger)^4 + D \min\{0, x\}^4.$$

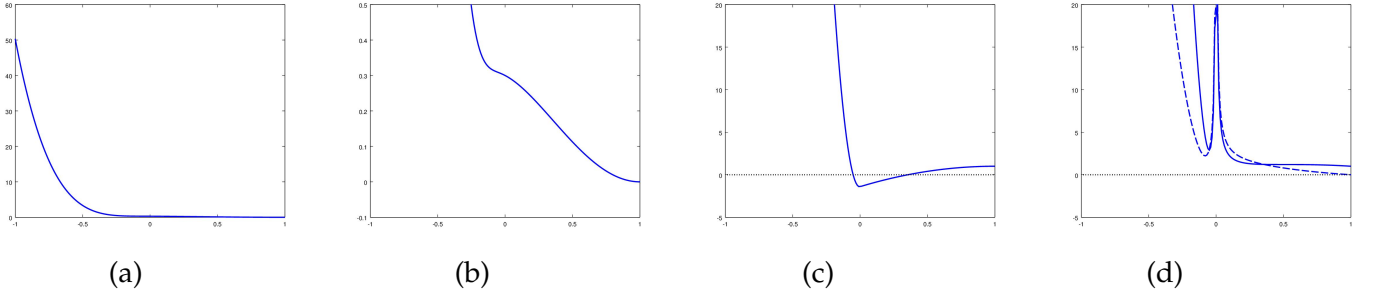


Figure 1: From left to right: (a) function r_1^δ , (b) detailed view of r_1^δ , (c) Hessian $r_1^{\delta''}$, (d) shifted Hessian $r_1^{\delta''} + \lambda_\rho$ (solid) and Lagrange multiplier λ_ρ (dashed)

It is easy to see that $x = 0$ cannot be a solution to (19), thus according to (22), (23), a solution x has to satisfy

$$\begin{aligned}
& x \in [-x^\dagger, x^\dagger] \setminus \{0\} \wedge \frac{-r_1^{\delta'}(x)}{x} \geq 0 \\
& \Leftrightarrow \left(-x^\dagger < x < 0 \wedge (x - x^\dagger)(1 - 4C(x - x^\dagger)^2) + 4Dx^3 \geq 0 \right) \\
& \quad \vee \left(x^\dagger > x > 0 \wedge (x - x^\dagger)(1 - 4C(x - x^\dagger)^2) \leq 0 \right) \\
& \Rightarrow \left(-x^\dagger < x < 0 \wedge 1 - 4C(x - x^\dagger)^2 \leq 0 \right) \vee \left(x^\dagger > x > 0 \wedge 1 - 4C(x - x^\dagger)^2 \geq 0 \right) \\
& \Leftrightarrow -x^\dagger < x < x^\dagger - \frac{1}{2\sqrt{C}} < 0 \vee x^\dagger > x > 0
\end{aligned}$$

where we have used the fact that $C < \frac{1}{4x^{\dagger 2}}$ in the last equivalence. Hence we get

$$\begin{aligned}
& x \in [-x^\dagger, x^\dagger] \setminus \{0\} \wedge \frac{-r_1^{\delta'}(x)}{x} \geq 0 \\
& \Rightarrow r_1^{\delta''}(x) + \frac{-r_1^{\delta'}(x)}{x} = 1 - 12C(x - x^\dagger)^2 + 12D \min\{0, x\}^2 + \frac{-r_1^{\delta'}(x)}{x} \\
& \quad \begin{cases} \geq 1 - 48Cx^{\dagger 2} + 12D(\frac{1}{2\sqrt{C}} - x^\dagger)^2 > 0 & \text{if } -x^\dagger < x < x^\dagger - \frac{1}{2\sqrt{C}} < 0 \\ = \frac{1}{x}(x^\dagger - 4C(2x^3 - 3x^2x^\dagger + x^{\dagger 3})) \geq \frac{x^\dagger}{x}(1 - 4Cx^{\dagger 2}) > 0 & \text{if } x^\dagger > x > 0 \end{cases}
\end{aligned}$$

where we have used the fact that $D > \frac{48Cx^{\dagger 2} - 1}{12(1/(2\sqrt{C}) - x^\dagger)^2}$ and $4Cx^{\dagger 2} < 1$. Thus condition (32) is satisfied although r_1^δ is nonconvex. An illustration of this example is provided in Figure 1.

Proposition 3.1. Let r_n^δ be twice Lipschitz continuously differentiable (30) and satisfy (32) at some minimizer $x_{n,\rho}^\delta$ of (19).

Then the iterates x^{k+1} defined as minimizers to (20) converge locally quadratically to $x_{n,\rho}^\delta$.

Proof. Let $x_{n,\rho}^\delta \in \operatorname{argmin}\{r_n^\delta(x_n) : x_n \in X_n, \|x_n\|_n^2 \leq \rho\}$. By finite dimensionality of the space X_n , condition (32) implies existence of $\alpha_n > 0$ such that

$$(33) \quad \forall w \in X_n : r_n^{\delta''}(x_{n,\rho}^\delta)w^2 - \frac{(r_n^{\delta'}(x_{n,\rho}^\delta), x_{n,\rho}^\delta)}{\rho} \|w\|_n^2 \geq \alpha_n \|w\|_n^2$$

Let the starting point $x^0 \in X_n$ be contained in some ϵ_n -neighborhood (wrt. $\|\cdot\|_n$) of $x_{n,\rho}^\delta$ with $\epsilon_n \in (0, \frac{\alpha_n}{3L})$. Using (33) with $w = x^{k+1} - x_{n,\rho}^\delta$ in (31) (recall that $r_n^{\delta''}(x_n^k) + \lambda^{k+1}I_n$ is positive semidefinite on all of X_n by (27)) for $k = 0$ by the triangle inequality yields

$$\alpha_n \|x_n^{k+1} - x_{n,\rho}^\delta\|_n^2 \leq L(2\|x_{n,\rho}^\delta - x_n^k\|_n^2 + \|x_{n,\rho}^\delta - x_n^k\|_n \|x_n^{k+1} - x_{n,\rho}^\delta\|_n) \|x_n^{k+1} - x_{n,\rho}^\delta\|_n,$$

hence

$$(34) \quad \begin{aligned} \|x_n^{k+1} - x_{n,\rho}^\delta\|_n &\leq \frac{2L}{\alpha_n - L\epsilon_n} \|x_{n,\rho}^\delta - x_n^k\|_n^2 \\ &\leq \frac{3L}{\alpha_n} \|x_{n,\rho}^\delta - x_n^k\|_n^2 \end{aligned}$$

$$(35) \quad \leq \frac{3L\epsilon_n}{\alpha_n} \|x_{n,\rho}^\delta - x_n^k\|_n$$

$$(36) \quad < \epsilon_n.$$

By an inductive argument using the same estimate for general k , the iterates remain in the ϵ_n -neighborhood of $x_{n,\rho}^\delta$ (cf. (36)) and satisfy a contraction (cf. (35)) as well as a quadratic convergence (cf. (34)) estimate. \square

We now consider these conditions in more detail in the special case of a Hilbert space norm

$$(37) \quad \mathcal{S}(y_1, y_2) = \frac{1}{2} \|y_1 - y_2\|^2$$

for measuring the data discrepancy. The optimality conditions (22)–(24) then in terms of the discretized forward operator F_n read as follows:

$$(38) \quad \begin{aligned} &\left\{ \begin{array}{l} \lambda_\rho = 0 \text{ and } \|x_{n,\rho}^\delta\|_n^2 \leq \rho \text{ and} \\ \forall w \in X_n : \langle F_n(x_{n,\rho}^\delta) - y^\delta, F_n'(x_{n,\rho}^\delta)w \rangle_n = 0 \text{ and} \\ \forall w \in X_n : \|F_n'(x_{n,\rho}^\delta)w\|_n^2 + \langle F_n(x_{n,\rho}^\delta) - y^\delta, F_n''(x_{n,\rho}^\delta)w^2 \rangle_n \geq 0 \end{array} \right. \\ &\text{or} \\ &\left\{ \begin{array}{l} \lambda_\rho = \frac{-\langle F_n(x_{n,\rho}^\delta) - y^\delta, F_n'(x_{n,\rho}^\delta)x_{n,\rho}^\delta \rangle_n}{\rho} > 0 \text{ and } \|x_{n,\rho}^\delta\|_n^2 = \rho \text{ and} \\ \forall w \in \{x_{n,\rho}^\delta\}^\perp : \langle F_n(x_{n,\rho}^\delta) - y^\delta, F_n'(x_{n,\rho}^\delta)w \rangle_n = 0 \text{ and} \\ \forall w \in \{x_{n,\rho}^\delta\}^\perp : \|F_n'(x_{n,\rho}^\delta)w\|_n^2 + \langle F_n(x_{n,\rho}^\delta) - y^\delta, F_n''(x_{n,\rho}^\delta)w^2 \rangle_n \geq 0 \end{array} \right. \end{aligned}$$

where in case (a) we use the fact that for all $w \in X_n$ either w or $-w$ is in the critical cone and in case (b) we have used the fact that for any $w \in X_n$

$$\langle F_n(x_{n,\rho}^\delta) - y^\delta, F'_n(x_{n,\rho}^\delta)w \rangle_n + \lambda_\rho \langle x_{n,\rho}^\delta, w \rangle_n = \langle F_n(x_{n,\rho}^\delta) - y^\delta, F'_n(x_{n,\rho}^\delta) \text{Proj}_{\{x_{n,\rho}^\delta\}^\perp} w \rangle_n$$

As a matter of fact this shows that the case (a) of a vanishing Lagrange multiplier is not really relevant here, since with $\hat{Y}_n := \overline{\mathcal{R}(F'_n(x_{n,\rho}^\delta))}$ it implies that $\text{Proj}_{\hat{Y}_n}(F_n(x_{n,\rho}^\delta) - y^\delta) = 0$ so that if $\overline{\bigcup_{n \in \mathbb{N}} \hat{Y}_n} = Y$ we get

$$r_n^\delta(x_{n,\rho}^\delta) = \frac{1}{2} \|F_n(x_{n,\rho}^\delta) - y^\delta\|_n^2 = \frac{1}{2} \|\text{Proj}_{\hat{Y}_n^\perp}(F_n(x_{n,\rho}^\delta) - y^\delta)\|_n^2 < \delta + \underline{\eta}_n^\delta - \eta_n^\delta,$$

where the last inequality holds for n sufficiently large, provided $\delta + \underline{\eta}_n^\delta - \eta_n^\delta > 0$, which is compatible with the assumptions made in Section 2. On the other hand, from (6) and (7) it follows that

$$r_n^\delta(x_{n,\rho(\delta)}^\delta) > \delta + \underline{\eta}_n^\delta - \eta_n^\delta$$

which by Lemma 2.1 implies $\rho > \rho(\delta)$.

The positivity condition in (32) becomes

$$(39) \quad \forall w \in X_n \setminus \{0\} : \|F'_n(x_{n,\rho}^\delta)w\|_n^2 + \langle F_n(x_{n,\rho}^\delta) - y^\delta, F''_n(x_{n,\rho}^\delta)w^2 \rangle_n + \lambda_\rho \|w\|^2 > 0.$$

This condition will indeed be satisfied for $\rho = \rho(\delta)$ for δ sufficiently small and n sufficiently large, e.g., in the situation of an estimate

$$\forall w \in X_n : \|F''_n(x_{n,\rho}^\delta)w^2\|_n \leq C \|F'_n(x_{n,\rho}^\delta)w\|_n^2$$

(which might be interpreted as a condition on the nonlinearity of the forward operator) holding with a constant C independent of n , since then for all $w \in X_n \setminus \{0\}$

$$\begin{aligned} & \|F'_n(x_{n,\rho}^\delta)w\|_n^2 + \langle F_n(x_{n,\rho}^\delta) - y^\delta, F''_n(x_{n,\rho}^\delta)w^2 \rangle_n + \lambda_\rho \|w\|^2 \\ & \geq \|F'_n(x_{n,\rho}^\delta)w\|_n^2 - (\tau\delta + \bar{\eta}_n^\delta) \|F''_n(x_{n,\rho}^\delta)w^2\|_n + \lambda_\rho \|w\|^2 \\ & \geq (1 - (\tau\delta + \bar{\eta}_n^\delta)C) \|F'_n(x_{n,\rho}^\delta)w\|_n^2 + \lambda_\rho \|w\|^2 > 0. \end{aligned}$$

However, Assumption (39) possibly remains valid also in case $x_{n,\rho}^\delta$ is still far away from x^\dagger since then the positive contribution of the (then typically larger) Lagrange multiplier takes effect: Note that both the residual $F_n(x_{n,\rho}^\delta) - y^\delta$ and the norm $\frac{1}{\sqrt{\rho}}$ of the ratio $\frac{x_{n,\rho}^\delta}{\rho}$ get larger for smaller $\rho < \|x^\dagger\|^2$.

Corollary 3.2. *Let Y be a Hilbert space, let $\mathcal{S}, \mathcal{S}_n$ be defined by the squared norm (37) and its finite dimensional approximation $\mathcal{S}_n(y_1, y_2) = \frac{1}{2} \|y_1 - y_2\|_n^2$, assume that F_n is twice Lipschitz continuously differentiable and that for some minimizer $x_{n,\rho}^\delta$ of (19) condition (39) holds.*

Then the iterates x_n^{k+1} defined as minimizers to (20) converge locally quadratically to $x_{n,\rho}^\delta$.

Remark 1. In view of the well-known equivalence between the Levenberg Marquardt method and the application of a trust region method to successive quadratic approximations of the nonlinear least squares cost function, there is an obvious relation to [3], still more, since we also use the discrepancy principle for choosing the trust region radius (as is done for the regularization parameter α in [3]). The main difference to the method described here, besides the somewhat more general data space setting, lies in the fact that we start from the nonlinear trust region problem and work with quadratic approximations of the cost function, that are possibly nonconvex. This is why the algorithm from [10] plays a key role here.

3.2 SOLVING THE NONCONVEX QUADRATIC TRUST REGION SUBPROBLEM

Discretization $x_n = \sum_{j=1}^n x^j \phi^j$ with a basis $\{\phi^1, \dots, \phi^n\}$ of X_n leads to an equivalent formulation of (20) as a constrained quadratic minimization problem

$$(40) \quad \min_{x \in \mathbb{R}^n} x^T A x - 2a^T x \text{ s.t. } x^T x \leq s^2$$

with a not necessarily positive semidefinite Hessian A . Such problems can be very efficiently solved by the method proposed in [10], which we briefly sketch in the following.

The key idea relies in the fact that solving (40) can be recast into a root finding problem for a parametrized eigenvalue problem. This can be motivated by the following chain of inequalities, where μ^* is the optimal value of (40) (see Section 2.2 in [10]).

$$\begin{aligned}
 \mu^* &= \min_{\|x\|=s, y_0^2=1} x^T A x - 2y_0 a^T x \\
 &= \max_t \min_{\|x\|=s, y_0^2=1} x^T A x - 2y_0 a^T x + t y_0^2 - t \\
 (41) \quad &\geq \max_t \min_{\|x\|^2 + y_0^2 = s^2 + 1} x^T A x - 2y_0 a^T x + t y_0^2 - t \\
 &\geq \max_{t, \lambda} \min_{x, y_0} x^T A x - 2y_0 a^T x + t y_0^2 - t + \lambda(\|x\|^2 + y_0^2 - s^2 - 1) \\
 &= \max_{r, \lambda} \min_{x, y_0} x^T A x - 2y_0 a^T x + r y_0^2 - r + \lambda(\|x\|^2 - s^2) \\
 &= \max_{\lambda} \left(\max_r \min_{x, y_0} x^T A x - 2y_0 a^T x + r y_0^2 - r + \lambda(\|x\|^2 - s^2) \right) \\
 &= \max_{\lambda} \min_x \min_{y_0^2=1} x^T A x - 2y_0 a^T x + \lambda(\|x\|^2 - s^2) = \mu^*
 \end{aligned}$$

Indeed, with $y = \begin{pmatrix} y_0 \\ x \end{pmatrix}$, the optimization problem on line (41) can be rewritten as

$$(42) \quad \max_t \min_{\|y\|^2 = s^2 + 1} y^T D(t) y - t = \max_t (s^2 + 1) \lambda_{\min}(D(t)) - t = \max_t k(t)$$

where

$$(43) \quad D(t) = \begin{pmatrix} t & -a \\ -a & A \end{pmatrix}$$

and $\lambda_{\min}(M)$ denotes the smallest (possibly negative) eigenvalue of some matrix M .

More precisely, it can be shown (Theorem 14 in [10]), that unless the so-called hard case occurs, for any $t \in \mathbb{R}$ and for any $\begin{pmatrix} y_0(t) \\ z(t) \end{pmatrix}$ being a normalized eigenvector corresponding to $\lambda_{\min}(D(t))$, the vector $x = \frac{1}{y_0(t)}z(t)$ is well-defined and solves (40) with $s = \frac{1-y_0(t)^2}{y_0(t)^2}$ (Here the “hard case” is the pathological situation of a being orthogonal to the eigenspace corresponding to $\lambda_{\min}(A)$.)

Based on this observation, it remains to iteratively find a root of the function $\psi(t) = \sqrt{s^2 + 1} - \frac{1}{y_0(t)}$ (that can be shown to be almost linear, nonincreasing and concave), or equivalently, to find a stationary point of the function k in (42), which can be done very efficiently using inverse interpolation, cf. [10].

The main computational effort of the resulting algorithm lies in the determination of an eigenvector corresponding to the smallest eigenvalue of $D(t)$ in each of these root finding iterations. For this purpose fast routines exist, (in our numerical tests we use the code from <http://www.math.uwaterloo.ca/~hwolkowi/henry/software/trustreg.d/> employing the Matlab routine `eigs` based on an Arnoldi method). Since this method only uses matrix vector products with A , it suffices to provide a routine for evaluating the linear operators $F'_n(x)$, $F''_n(x)^*(F_n(x) - y^\delta)$ in a given direction d , which particularly makes sense for F_n being the (discretization of a) forward operator for some inverse problem, e.g., some parameter identification problem in a PDE. Namely, in that case these actions just correspond to solving the underlying linearized PDE model with some inhomogeneity involving d .

3.3 NEWTON'S METHOD FOR COMPUTING $\rho(\delta)$

In view of the discrepancy principle (7) for choosing $\rho = \rho(\delta)$, we have to approximate a root of the one-dimensional function ϕ defined by

$$(44) \quad \phi(\rho) = r_n^\delta(x_{n,\rho}^\delta) - r_d \quad \text{where } r_d = \frac{(\tau + 1)\delta + \bar{\eta}_n^\delta + \underline{\eta}_n^\delta}{2},$$

which by Lemma 2.1 is monotonically decreasing. For this purpose, Newton's method is known to converge globally and quadratically, provided ϕ is twice continuously differentiable. As a matter of fact, in the generic case of $x_{n,\rho}^\delta$ lying on the boundary of the feasible set, the derivative of ϕ is just the Lagrange multiplier for (19), since by the complementarity condition in (23)

$$\begin{aligned} \phi'(\rho) &= \frac{d}{d\rho} r_n^\delta(x_{n,\rho}^\delta) = \frac{d}{d\rho} \left(r_n^\delta(x_{n,\rho}^\delta) + \lambda_\rho (\|x_{n,\rho}^\delta\|^2 - \rho) \right) \\ &= r_n^{\delta'}(x_{n,\rho}^\delta) \frac{dx_{n,\rho}^\delta}{d\rho} + \frac{d\lambda_\rho}{d\rho} (\|x_{n,\rho}^\delta\|^2 - \rho) + \lambda_\rho \langle x_{n,\rho}^\delta, \frac{dx_{n,\rho}^\delta}{d\rho} \rangle_n - \lambda_\rho \\ &= -\lambda_\rho \end{aligned}$$

by (22) and $\|x_{n,\rho}^\delta\|^2 = \rho$. A similar observation has already been made for the derivative of the cost functional with respect to the regularization parameter in Tikhonov regularization cf. [7].

3.4 ALGORITHM

Altogether we arrive at the following nested iteration.

Algorithm 1. .

- 1: Given $\delta > 0$, choose $\tau > 1$, n , η_n^δ , $\bar{\eta}_n^\delta$, η_n^δ such that (8), (9), (18) are satisfied, $\rho_0 = 0$, $\rho_1 > 0$;
- 2: Set $l = 1$, $\hat{x}_{n,\rho_0}^\delta = 0$, $\hat{x}_{n,\rho_1}^\delta = 0$;
- 3: **while** $r_n^\delta(\hat{x}_{n,\rho_l}^\delta) > \tau\delta + \bar{\eta}_n^\delta$ **do**
- 4: $k = 0$, $x_n^0 = \hat{x}_{n,\rho_{l-1}}^\delta$
- 5: **while** $r_n^\delta(x_n^k) > \min_{x_n \in X_n} \{r_n^\delta(x_n) : \|x_n\|_n^2 \leq \rho_l\} + \eta_n^\delta$ **do**
- 6: Compute solution x_n^{k+1} and Lagrange multiplier λ^{k+1} to (20), (21), $\rho = \rho_l$
- 7: Set $k = k + 1$
- 8: Set $\hat{x}_{n,\rho_l} = x_n^k$, $\lambda_{\rho_l} = \lambda^k$
- 9: Set $\rho_{l+1} = \rho_l + \frac{r_n^\delta(\hat{x}_{n,\rho_l}^\delta) - r_d}{\lambda_{\rho_l}}$ with r_d as in (44)
- 10: Set $l = l + 1$

4 NUMERICAL TESTS

To illustrate performance of Algorithm 1 we make use of an implementation of the method described in [10] available on the web page of one of the authors <http://www.math.uwaterloo.ca/~hwolkowi/henry/software/trustreg.d/> and consider the nonlinear integral equation

$$y(t) = \int_0^t (x(s) + 10x(s)^3) ds \quad t \in (0, 1)$$

with $X = Y = L^2(0, 1)$ and $\mathcal{S}(y_1, y_2) = \frac{1}{2}\|y_1 - y_2\|_{L^2}^2$. The integral equation is discretized with a composite trapezoidal rule on an equidistant grid with 100 breakpoints.

Figure 2 shows the noisy data as well as the exact and reconstructed solutions with different noise levels. In figure 3 we plot error and residual, as well as the number of solved nonconvex quadratic subproblems over the radius, for noise levels 1 and 3 per cent, during the iteration over ρ . Figure 4 illustrates that condition (32) remains valid throughout the iteration over ρ , while several nonconvex trust region subproblems have to be solved (the number being smaller for large δ since less iterations are carried out in that case).

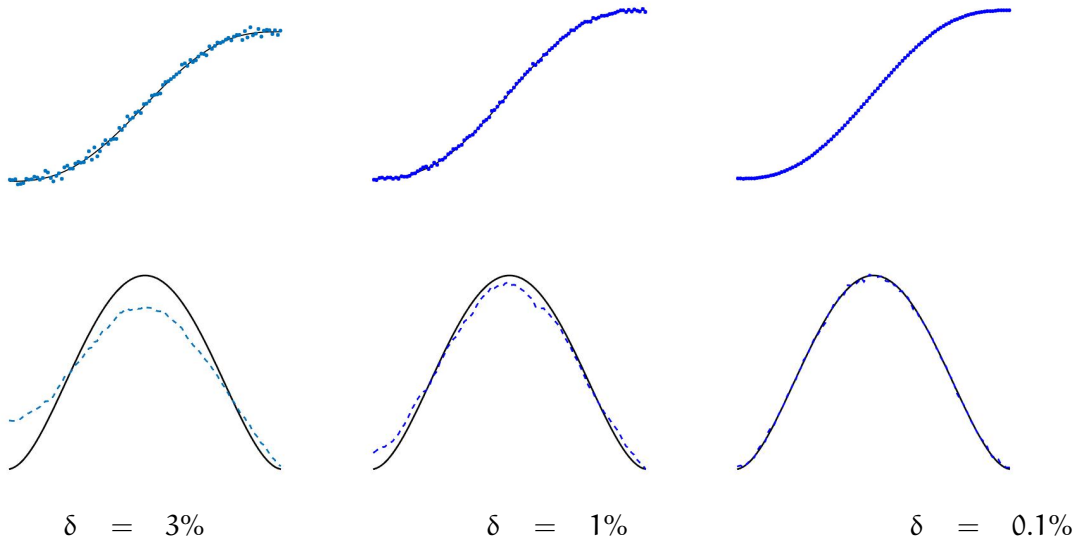


Figure 2: Noisy data (top row) and reconstructions (bottom row) for $\delta = 3\%$ (left column), $\delta = 1\%$ (middle column), $\delta = 0.1\%$ (right column)

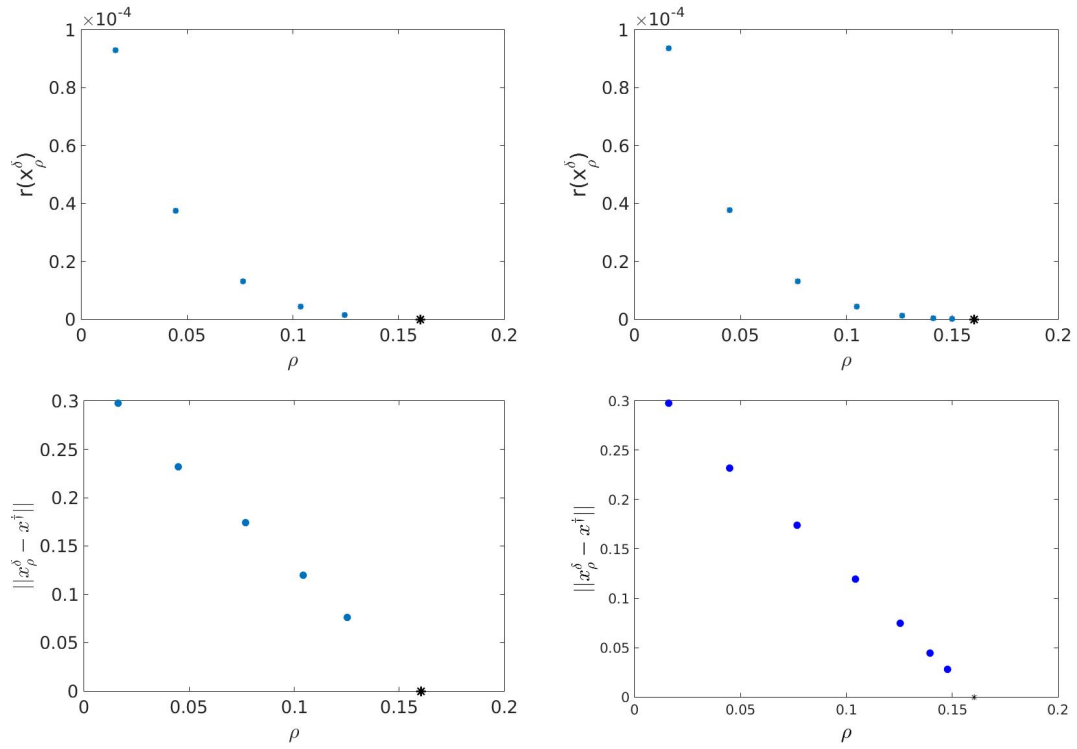


Figure 3: Residuals (top row) and errors (bottom row) for $\delta = 3\%$ (left column) and $\delta = 1\%$ (right column).

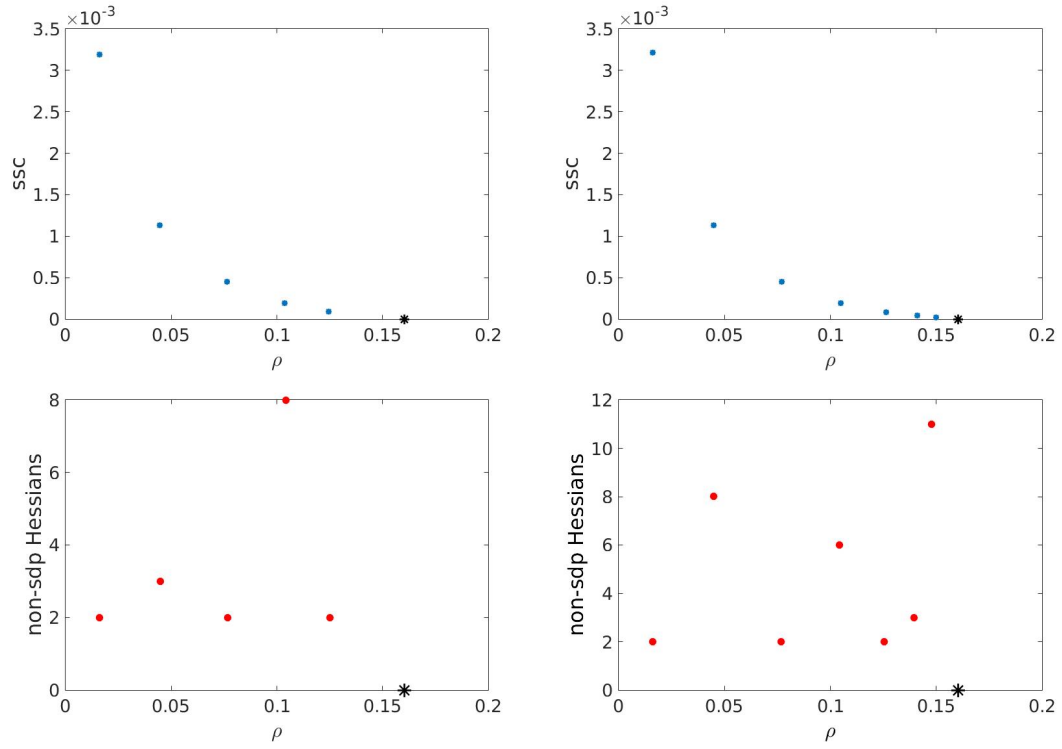


Figure 4: Smallest eigenvalue of Hessian of Lagrangian (top row) and number of non-convex quadratic subproblems (bottom row) for $\delta = 3\%$ (left column) and $\delta = 1\%$ (right column).

REFERENCES

- [1] C. CLASON, B. KALTENBACHER, AND A. KLASSEN, *On convergence and convergence rates for Ivanov and Morozov regularization*, (2015). in preparation.
- [2] O. GRODZEVICH AND H. WOLKOWICZ, *Regularization using a parameterized trust region subproblem*, Math. Program., Ser. B, 116 (2009), pp. 193–220.
- [3] M. HANKE, *A regularization Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems, 13 (1997), pp. 79–95.
- [4] V. K. IVANOV, *On linear problems which are not well-posed*, Dokl. Akad. Nauk SSSR, 145 (1962), pp. 270–272.
- [5] ———, *On ill-posed problems*, Mat. Sb. (N.S.), 61 (103) (1963), pp. 211–223.
- [6] V. K. IVANOV, V. V. VASIN, AND V. P. TANANA, *Theory of Linear Ill-posed Problems and Its Applications*, Inverse and ill-posed problems series, VSP, 2002.
- [7] B. KALTENBACHER, A. KIRCHNER, AND B. VEXLER, *Adaptive discretizations for the choice of a Tikhonov regularization parameter in nonlinear inverse problems*, Inverse Problems, 27 (2011), p. 125008.
- [8] D. LORENZ AND N. WORLICZEK, *Necessary conditions for variational regularization schemes*, Inverse Problems, 29 (2013), p. 075016.
- [9] A. NEUBAUER AND R. RAMLAU, *On convergence rates for quasi-solutions of ill-posed problems.*, ETNA, Electron. Trans. Numer. Anal., 41 (2014), pp. 81–92.
- [10] F. RENDL AND H. WOLKOWICZ, *A semidefinite framework for trust region subproblems with applications to large scale minimization*, Math. Program., 77 (1997), pp. 273–299.
- [11] T. I. SEIDMAN AND C. R. VOGEL, *Well posedness and convergence of some regularisation methods for non-linear ill posed problems*, Inverse Problems, 5 (1989), p. 227.
- [12] D. SORENSEN, *Newton’s method with a model trust region modification*, SIAM J. Numer. Anal., 19 (1982), pp. 409–426.
- [13] C. R. VOGEL, *A constrained least squares regularization method for nonlinear ill-posed problems*, SIAM Journal on Control and Optimization, 28 (1990), pp. 34–49.